Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 48 (2015) 84 – 89

Procedia Computer Science

International Conference on Intelligent Computing, Communication & Convergence

(ICCC-2015)

Conference Organized by Interscience Institute of Management and Technology,

Bhubaneswar, Odisha, India

Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction

P.G Preethi[a],V. Uma[b] ,Ajit kumar[c]

[abc]*Department Of Computer Science, School of Engineering & Technology,Pondicherry University, Puducherry- 605014,India*

Abstract

Sentiment analysis or opinion mining is the process of computationally identifying and categorizing opinions expressed in a piece of text, in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral. It is one of the most active research areas in natural language processing and text mining in recent years. A detailed study of the two concepts (1) Temporal sentiment analysis (2) Sentiment causal relation is presented in this paper. Temporal sentiment analysis is useful for summarizing the events based on sentiment and time. Causal relation is useful for identifying cause and effect of events and is also useful for event prediction. These two concepts when combined result in a better event prediction model that can predict the time period between the events and sentiment of upcoming events. The proposed work introduces a generalized prediction model based on temporal sentiment analysis of tweet to identify the causal relation between the events which can be used to predict the event sentiment and duration between the events. The proposed method is to be evaluated using the performance measures precision and recall. The accuracy of causal rule prediction is evaluated using parameters Mean Absolute Error (MAE) and the Root Mean Squared error (RMSE).
Keywords:Social media ;Sentiment analysis;Temporal sentiment analysis;Causal rule;Prediction

1. Introduction

Social media is a collection of internet based application that allows the users to create, share or exchange information [1]. Most common examples for social media are Social Network Sites (SNS), blogs, micro blogs etc. Social media sentiment analysis is a social media mining technique that involves constructing systems to collect and analyze opinions about different products that appear in blog posts, comments, reviews or tweets.

Sentimental analysis or opinion mining is the field that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions towards some entities such as products, services, organizations, individuals, issues, events, topics etc. The terms sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining come under the same category commonly known as sentiment analysis or opinion mining [2]. Recent research indicates that analysis of the online texts can be useful for

Peer-review under responsibility of scientific committee of International Conference on Computer, Communication and Convergence (ICCC 2015)
doi:10.1016/j.procs.2015.04.154

trend or event prediction. In [3] political orientations are predicted from the comments and reviews and [4] movie box office success is predicted on the basis of sentimental analysis. In [5] news sentiment or twitter mood is used to predict stock markets movement.

Events are usually anchored to temporal expressions and these temporal attributes of the events are useful for identifying the temporal relation between the events and ordering of the events. These temporal attributes are also useful in better presentation of news and better prediction of events. A more complex type of relationship between events is causality. Identifying the causal relation between events is an important step in predicting occurrence of future events [6].

The proposed work introduces a prediction model based on temporal sentiment analysis to identify the causal relation between the events and uses it to predict the event sentiment and duration between the events. The proposed work use Support Vector Machine (SVM) for sentiment classification and the causal relation is found using support and confidence. The remainder of the paper is structured as follows. Section 2 briefs the survey on sentiment analysis. Section 3 discuss the relation between sentiment analysis and temporal relation Section 4 discusses about causal rule detection and sentiment analysis Section 5 overview of proposed system and in Section 6 conclusion and future work is provided.

2. Literature Survey

This section presents the survey on various methods for sentiment analysis, sentiment classification, temporal sentiment analysis and causal rule detection

2.1 Sentiment Analysis

Sentimental analysis can often be conducted in three levels known as Document Level, Sentence Level and Entity and Aspect Level. Document Level sentiment analysis classifies the entire document into either positive or negative [7]. Sentence level classification classifies the sentence into positive, negative or neutral category. In [8] polarity prediction model for sentence level sentiment classification was introduced. Entity and Aspect level also known as feature level sentimental analysis gives the summary about which feature of a product does user like or dislike [9].

2.2 Sentiment Classification Method

Different approaches or methods such as supervised learning, unsupervised learning and semi-supervised learning can be used for sentiment classification. Vaithyanathan et al.[7] applied supervised machine learning methods (Naïve Bayes, maximum entropy classification and support vector machine) for sentiment classification and evaluated its effectiveness in movie domain and concluded that SVM method showed better performance compared to other methods while Naïve Bayes showed the worst performance. Mullen et al. [10] proposed a SVM based sentiment classification method that assigned values to selected phrases and words, and used a technique for bringing them together to create a model for classification of texts. Qiang Ye et al. [11] used a supervised method in traveler review sites and found the sentiment based on user reviews and also proved that the SVM outperformed Naïve Bayes approach. Deng et al. [12] introduced a new term weight method based on two factors; first one being the importance of the document and the second one, the importance of the term for expressing the sentiment. Advantage of this method is that, it can make full use of the available labeling information to assign appropriate weights to terms. In most of the sentiment analysis works based on supervised learning, SVM shows high performance and accuracy

The proposed work intents to use twitter data for sentiment analysis. SVM has following advantages(1) it is highly robust to over fitting (2) it can handle large feature spaces [13] [14], (3) it is extremely efficient in learning sentiments from large twitter dataset and also supports prediction of information So in proposed work SVM based sentiment analysis prediction is done.

3 Sentiment Analysis and Temporal Relation

Mishne and Rijke proposed [15] a system called Moodview for tracking and analyzing the mood of bloggers worldwide. Mood view can analyze the temporal change of sentiment. Fukuhara et al.[16] proposed a method for analyzing temporal trends of sentiments and topics from documents with timestamps. Das et al. [17] proposed a method for finding the contribution of sentiments in determining the event-event relations from text. Usually event sentiment over time is calculated based on the web content like tweets, blogs, normal news article sites etc. These methods can easily summarize the events based on the time and overall sentiment. The proposed work uses both sentiment analysis and temporal relations for predicting the events and the average time period between two events using twitter blog.

4 Sentiment Analysis And Causal Relation

Causality (also referred to as causation) is the relation between an event (the cause) and a second event (the effect), where the second event is understood as a physical consequence of the first. The causal relation shows how variations in one variable can cause changes in the other variable. Therefore causality is more useful for prediction and reasoning. The contributions done in the field of sentiment analysis and causal relation detection is presented in Table 1. The comparison of the existing works with the proposed work is also given in the table.

5 Proposed Work

A brief idea about the proposed work is illustrated in this section. The proposed work mainly consist of four phases. The Fig.1 shows the detailed architecture of the proposed system.

TABLE.1: Comparison of works done on Sentiment Analysis and Causal rule extraction

Sl No	Title	Contribution	Method Used	Causal Relation identification	Limitation
1	Topic Sentiment Change Analysis [18]	Topic Sentiment Change Analysis	1.Sentiment analysis 2.Time base analysis 3.Causal rule detection	Identifying cause for sentiment change	No prediction
2	Predictive Sentiment Analysis of Tweets: A Stock Market Application[20]	Twitter feed can forecast stock market movement	1.Sentiment analysis 2.Time base analysis 3.Causalrule detection 4.Prediction	1.Causal identification using positive sentiment probability 2. Twitter sentiment causes stock change.	1.Time duration between events is not predicted 2. Concentrates only on finance and stock market.
3	Extracting Temporal and Causal relations between Events[6]	Temporal causality identification	1.Time based analysis 2.Causal rule detection	Causal relations are identified based on annotation framework for events and temporal relations, namely TimeML	1.No sentiment evaluation 2.No prediction
4	Facebook's daily sentiment and international stock markets [19]	Evaluating the daily sentiment and trading behavior using face book gross national happiness index	1.Sentiment analysis 2.Time Based analysis 3.Causal rule identification	1. Causal relation between Sunday's sentiments and Monday's stock change	1. No prediction of time duration between two events is done. 2.Suitable for stock market domain
5	Stream-based active learning for sentiment analysis in the financial domain. [21]	Active learning approach for sentiment analysis of tweet streams in the stock market domain	1.Sentiment analysis 2.Time based analysis 3.Causal rule detection 4.Prediction	1.Causal identification using positive sentiment probability 2. Twitter sentiment cause stock variation. 3. Correlating two domain using causal rule.	1.No prediction of time duration between two events is done 2 Concentrates on finance or stock market field only
6	Sentimental causal rule discovery from Twitter[22]	Sentimental causal rule discovery using twitter	1.Sentiment analysis 2.Causal rule detection	Combines sentiment analysis and causal rule and forms sentiment causal rule	1.No temporal factor considered 2.No prediction
7	Proposed work	A generalized approach that uses sentiment causal rule for event sentiment prediction	1.Sentiment analysis 2.Time based Analysis 3.Causal rule detection 4. Prediction of event sentiment and time period between two events.	Uses support and confidence for causal rule identification and uses the causal rule for future events prediction	Approximate prediction of time period between the events.

Step 1 Keyword Extraction: The first step of the proposed work is extracting aspect keywords from tweets with in a time period. These aspect keywords are used for causal rule detection After extracting the aspect keywords the next step is identifying the sentiment of the aspect keywords. The Alchemy API is used for the Keyword extraction and the Sentiment Analysis phase.

Step 2 Sentiment Analysis: The main objective of this phase is determining the sentiment of the aspect keyword. Identifying the polarity of a keyword is a difficult task because positive word may be negative in some situation and negative word maybe considered as positive in other situation. Since SVM shows high performance, it is used for sentiment classification. SVM classification is based on features. Feature selection is to be optimized so that classification error is reduced. To achieve higher classification accuracy 10-fold cross validation is to be performed**.** The method used for sentiment analysis is to be evaluated using the performance measure precision and recall. Precision and Recall can be calculated [11] using (1) (2) (3) and (4)

$$\text{Precision (pos)} = \frac{a}{a+b} \quad (1)$$

$$\text{Precision (neg)} = \frac{d}{d+c} \quad (2)$$

$$\text{Recall (pos)} = \frac{a}{a+c} \quad (3)$$

$$\text{Recall (neg)} = \frac{d}{d+b} \quad (4)$$

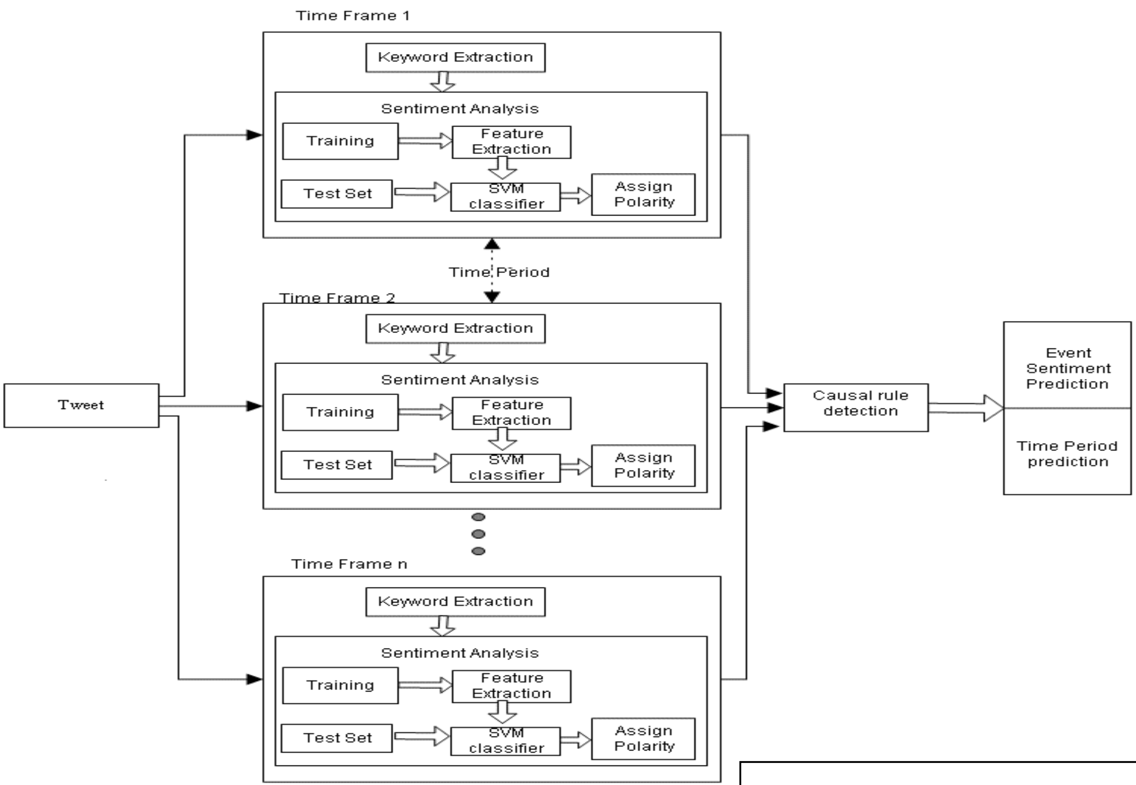

Fig. 1 Block Diagram of proposed work

where a, b are true positive and negative review for positive polarity prediction and c,d are true positive and negative
review for negative polarity prediction.

Step 3 Causal rule detection: The next step is identifying the causal rule between aspect keywords. The causal rule is detected by Support and Confidence [22] using (5) and (6)

$$\text{Supp}\,(I_1\,,\ I_2\ \rightarrow\ I_3)\ =\ \frac{|\omega_{12}|}{|T|} \tag{5}$$

$$\text{Confi}\,(I_1\,,\ I_2\ \rightarrow\ I_3)\ =\frac{|\omega_{123}|}{|\omega_{12}|} \tag{6}$$

where is the set of all tweets and ω_{12} is a subset of transactions including I_1 and I_2 and ω_{123} is a subset of transactions including and I_1 ,I_2 and I_3

Step 4 Prediction: The final step is prediction based on causal rule identified in the previous step. Using time based analysis of tweets and causal rule the prediction of upcoming events is done. Accuracy of prediction is evaluated using the parameters Mean Absolute Error (MAE) and the Root Mean Squared error (RMSE) [23] by (7) and (8) MAE and RMSE measures the deviation between the true rating and predicted value

$$\text{MAE}=\frac{\sum_i |p_i - r_i|}{n} \tag{7}$$

$$\text{RMSE}=\sqrt{\frac{\sum_i (p_i - r_i)^2}{n}} \tag{8}$$

where r_i is the true value and p_i is the predicted value of i^{th} tweet The experiments are to be conducted to minimize the objective functions specified in Eqn. 7 & 8.

.

6 Conclusion

In this proposed work Sentiment analysis is done on tweet data and the people’s attitude or sentiment towards a topic or incident is identified. Analyzing the people’s opinion in different time period is useful for causal rules detection .The causal rules identified are useful for event prediction. Time based analysis of tweets help in identifying the possible time period between the predicted events. Future work is to implement the system, evaluate using the given performance measures and prove the efficacy over the existing system.

References

1. Kaplan, A.M., Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of Social Media. Bus. Horiz. , 59–68. doi:10.1016/j.bushor.2009.09.003
2. Liu, B., 2012. Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. 5, 1–167.
3. Park, S., Ko, M., Kim, J., Liu, Y., Song, J., 2011.The politics of comments: predicting political orientation of news stories with commenters’ sentiment patterns, in: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work. ACM, pp. 113–122.
4. Schuller, B., Knaup, T., 2011. Learning and knowledge-based sentiment analysis in movie review key excerpts, in: Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues. Springer, pp. 448–472.
5. Li, X., Xie, H., Chen, L., Wang, J., Deng, X., 2014. News impact on stock price return via sentiment analysis. Knowl.-Based Syst. 69, 14–23. doi:10.1016/j.knosys.2014.04.022
6. Mirza, P., 2014. Extracting Temporal and Causal Relations between Events. ACL 2014 10.

7. Pang, B., Lee, L., Vaithyanathan, S., 2002. Thumbs up?: sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10. Association for Computational Linguistics, pp. 79–86.
8. Tan, L.K.-W., Na, J.-C., Theng, Y.-L., Chang, K., 2011. Sentence-level sentiment polarity classification using a linguistic approach, in: Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation. Springer, pp. 77–87.
9. Wiebe, J.M., Bruce, R.F., O'Hara, T.P., 1999. Development and use of a gold-standard data set for subjectivity classifications, in: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics. Association for Computational Linguistics, pp. 246–253.
10. Mullen, T., Collier, N., 2004. Sentiment Analysis using Support Vector Machines with Diverse Information Sources., in: EMNLP. pp. 412–418.
11. Ye, Q., Zhang, Z., Law, R., 2009. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Syst. Appl. 36, 6527–6535. doi:10.1016/j.eswa.2008.07.035
12. Deng, Z.-H., Luo, K.-H., Yu, H.-L., 2014. A study of supervised term weighting scheme for sentiment analysis. Expert Syst. Appl. 41, 3506–3513. doi:10.1016/j.eswa.2013.10.056
13. Joachims.T, Text categorization with support vector machines: learning with many relevant features, in: Proceedings of the European Conference on Machine Learning, 1998, pp. 137–142.
14. Sebastiani .F, "Machine learning in automated text categorization." ACM computing surveys (CSUR) 34, no. 1 (2002): pp 1-4
15. Mishne, G., De Rijke, M., 2006. MoodViews: Tools for Blog Mood Analysis. in: AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs. pp. 153–154.
16. Fukuhara, T., Nakagawa, H., Nishida, T., 2007. Understanding Sentiment of People from News Articles: Temporal Sentiment Analysis of Social Events. in: ICWSM.
17. Das, D., Kolya, A.K., Ekbal, A., Bandyopadhyay, S., 2011. Temporal Analysis of Sentiment Events– A Visual Realization and Tracking, in: Computational Linguistics and Intelligent Text Processing. Springer, pp. 417–428.
18. Jiang, Y., Meng, W., Yu, C., 2011. Topic sentiment change analysis, in: Machine Learning and Data Mining in Pattern Recognition. Springer, pp. 443–457.
19. Siganos, A., Vagenas-Nanos, E., Verwijmeren, P., 2014. Facebook's daily sentiment and international stock markets. J. Econ. Behav. Organ. doi:10.1016/j.jebo.2014.06.004
20. Smailović, J., Grčar, M., Lavrač, N., Žnidaršič, M., 2013. Predictive sentiment analysis of tweets: A stock market application, in: Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data. Springer, pp. 77–88.
21. Smailović, J., Grčar, M., Lavrač, N., Žnidaršič, M., 2014. Stream-based active learning for sentiment analysis in the financial domain. Inf. Sci. 285, 181–203. doi:10.1016/j.ins.2014.04.034
22. Dehkharghani, R., Mercan, H., Javeed, A., Saygin, Y., 2014. Sentimental causal rule discovery from Twitter. Expert Syst. Appl. 41, 4950–4958. doi:10.1016/j.eswa.2014.02.024
23. Li, F., Wang, S., Liu, S., Zhang, M., 2014. SUIT: A Supervised User-Item Based Topic Model for Sentiment Analysis, in: Twenty-Eighth AAAI Conference on Artificial Intelligence.